Bayes Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Simple example

Example: What is the probability that a fair dice rolled a 1, given that it rolled an odd number?

By Bayes theorem, using *A*=1 and *B*=odd, we have

$$P(1|odd) = \frac{P(1 \cap odd)}{P(odd)} = \frac{P(1)}{P(odd)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Frequentists can interpret this probability under an experiment where a dice is rolled repeatedly, but even-numbered rolls are discarded.

Bayesian use of Bayes

$P(A|B)=\frac{P(A\cap B)}{P(B)}$ appears in the form $\pi(\theta\,|\,y)=\frac{f(y,\theta)}{f(y)}$

- θ denotes the unobservable quantities.
  - Often just the parameters, but in some models will include random effects, process errors, predictions etc.
- *y* denotes the data.
- $f(y,\theta)$ is the joint density of unobservables and data.
- $f(y)$ is the (marginal) density of the data.
- $\pi(\theta\,|\,y)$ is the posterior density of the unobservables given the data.

Priors and likelihoods

The fundamental theorem of probability:

$$f(y,\theta)=\pi(\theta)\,f(y\,|\,\theta)$$

where

- The prior density of θ, is denoted $\pi(\theta)$.
- $f(y\,|\theta)$ is the model for the data (e.g., regression, ANOVA, binomial, Poisson, etc.). This gives the density function for the data. This is also known as the likelihood function.

Thus, the posterior can be obtained as

$$\pi(\theta\,|\,y)=\frac{f(y,\theta)}{f(y)}=\frac{f(y\,|\,\theta)\pi(\theta)}{f(y)}$$

Priors

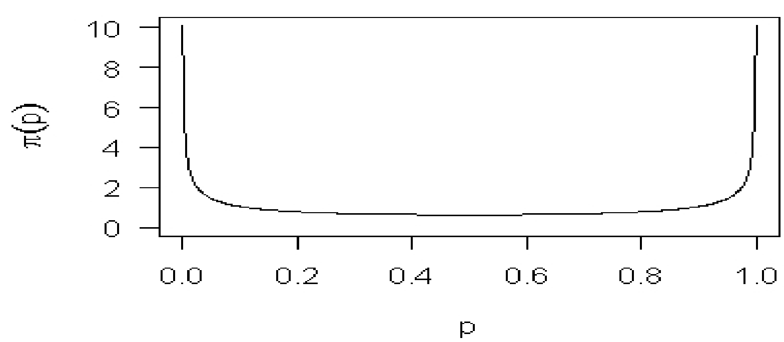
Prior information must be specified for every parameter in the model.

$$\pi(\theta) = \pi(\theta_1, \theta_2, ..., \theta_p)$$

where $\pi(\theta)$ is a joint density function.

Example: Binomial Priors

If we observe Binomial(n,p) data, the "reference" prior $\pi(p)$ is:

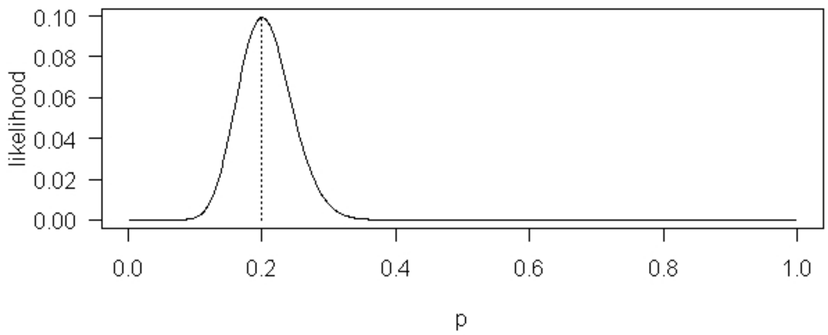


$$\pi(p) = \frac{1}{\pi\sqrt{p(1-p)}}$$

Example: Binomial likelihood

If we observe y=20 successes from 100 trials then the likelihood for *p*=(prob of success) is

$$f(20 \mid p) = \binom{100}{20} p^{20} (1-p)^{80}$$



Model likelihood *f(y)*

Bayes formula: $\pi(\theta \mid y) = \frac{f(y,\theta)}{f(y)} = \frac{f(y \mid \theta)\pi(\theta)}{f(y)}$

The denominator of Bayes formula, *f(y)*, is often called the model likelihood or marginal likelihood.

- *f(y)* acts solely as a "normalizing constant" (it does not involve θ) and can usually be ignored when working with $\pi(\theta \mid y)$.
- *f(y)* is important for model comparison.

The formula for f(y) is

$$f(y) = \int f(y,\theta)d\theta = \int f(y \mid \theta)\pi(\theta)d\theta$$

Example 1: Binomial data

Observe Y=20 from a Binomial(100,*p*) experiment.

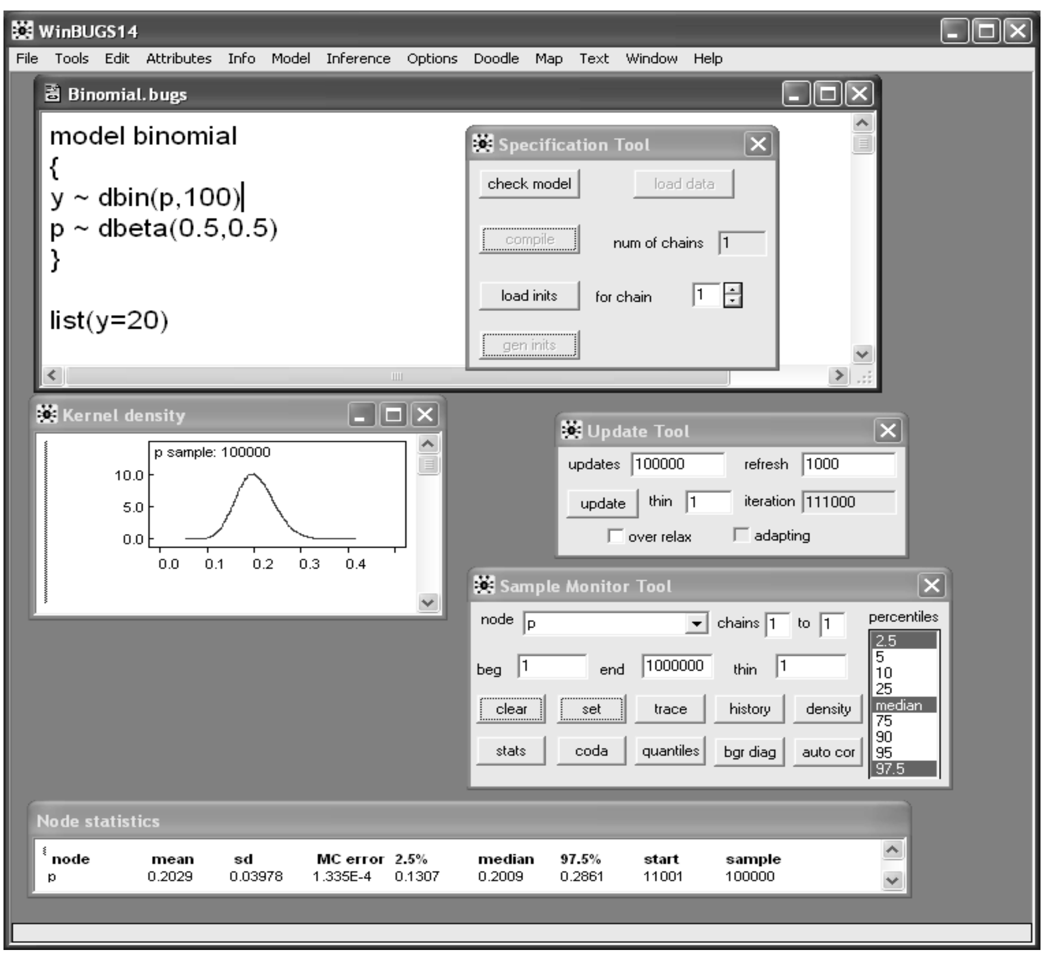
We need to calculate $\pi(p \mid y) = \dfrac{f(y \mid p)\pi(p)}{f(y)}$

where $\pi(p) \propto p^{-1/2}(1-p)^{-1/2}$, $0<p<1$

and $f(y,p) \propto p^{20}(1-p)^{80}$

So, $\pi(p|y) \propto p^{19.5}(1-p)^{79.5}$, $0<p<1$

The trick is to recognize that $\pi(p|y)$ is the density function of a Beta(20.5,80.5) distribution.



Example 2: IID Normal data

Here, the data (Y_1, Y_2, Y_3, …, Y_n) are independent and identically distributed as $N(\mu,\sigma^2)$. [This is conditional on μ and σ^2.]

To keep the calculus manageable, it is assumed that σ^2 is known, so that μ is the only unknown.

The prior on μ is $N(\nu,\phi^2)$.

Algebra →


Denote $\mathbf{y} = y_1, y_2, \ldots, y_n$.
It makes life easier to recognize that $f(\mathbf{y}|\mu) \propto f(\bar{y}|\mu)$
where $\bar{y}$ is the sample mean.
Then we have

$$\pi(\mu|\mathbf{y}) = \frac{\pi(\mu) f(\mathbf{y}|\mu)}{f(\mathbf{y})} \propto \pi(\mu) f(\bar{y}|\mu) . \qquad (1)$$

Now,

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}\phi} e^{-(\mu-\nu)^2/2\phi^2}$$

$$f(\bar{y}|\mu) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(\bar{y}-\mu)^2/2\sigma^2}$$

and so

$$\begin{aligned} \pi(\mu|\mathbf{y}) &\propto e^{-(\mu-\nu)^2/2\phi^2 - n(\bar{y}-\mu)^2/2\sigma^2} \\ &\propto \text{.....algebra} \\ &\propto \text{.....more algebra} \\ &= e^{-\left(\mu - \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\phi^2}\nu}{\frac{n}{\sigma^2} + \frac{1}{\phi^2}}\right)^2 / 2\left(\frac{n}{\sigma^2} + \frac{1}{\phi^2}\right)^{-1}} \end{aligned}$$

1

Example concluded

The formula for $\pi(\mu \mid y)$ corresponds to a normal density. Specifically,

$$\mu \mid y \sim N\left(\nu^*, \phi^{2^*}\right)$$

where

$$\nu^* = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\phi^2}\nu}{\frac{n}{\sigma^2} + \frac{1}{\phi^2}}$$

$$\phi^{2^*} = \left(\frac{n}{\sigma^2} + \frac{1}{\phi^2}\right)^{-1}$$

Note that the posterior mean is a weighted average of the prior mean and sample mean.

IID example in WinBUGS

The model is

$$Y_i \mid \mu, \sigma^2 \sim N(\mu,\sigma^2),\ i=1,\ldots n, \quad \text{and} \quad \mu \sim N(\nu,\phi^2).$$

```
model IIDNormal
{
  for(i in 1:n) { y[i] ~ dnorm(mu,prec.y) }
  prec.y <- 1/sigma2
  mu ~ dnorm(nu,prec.mu)
  prec.mu <- 1/phi2
}
...plus a few details we shall see later
```


The general case

$$\pi(\theta \mid y) = \frac{f(y,\theta)}{f(y)} = \frac{f(y \mid \theta)\pi(\theta)}{f(y)}$$

In general, the calculus required to work with the above formula is formidable (to say the least). In special cases, if the prior is chosen to "match" the likelihood, then the calculus is manageable. These are known as conjugate priors.

Until the advent of MCMC, the computational difficulties were a major disadvantage of Bayesian modeling. Now, it is the other way around, with MCMC permitting the easy fitting of models that may be intractable to frequentist statistics!